

## **Hedging one's bets too much? A reply to Levelt (2002)**

F.-Xavier Alario

*Cognitive Neuroimaging Unit, INSERM U-562, Orsay, France*

Albert Costa

*SISSA, Trieste, Italy, and Universitat de Barcelona, Spain*

Alfonso Caramazza

*Harvard University, Cambridge, MA, USA*

Levelt (2002) challenges the theoretical motivation and the interpretation of the data reported in Alario, Costa, and Caramazza's (2002) study on utterance planning during speech production. In this response we argue against these criticisms. First, we show that the hypotheses entertained in our research about the scope of phonological encoding are well motivated in the context of current theories of speech production. Second, we argue that although alternative interpretations of the frequency effect we report are logically possible, the available empirical evidence makes them very unlikely.

Experimental research in the field of speech production has traditionally investigated the processes of single word production (Levelt, Roelofs, & Meyer, 1999). In recent years, however, there has been a growing interest in the production of multi-word utterances: A number of studies have addressed this issue with the experimental methods originally developed for the study of single word production (e.g., the picture–word interference paradigm; Briggs & Underwood, 1982; Lupker, 1979, 1982). The use of

---

Requests for reprints should be sent to Alfonso Caramazza, Department of Psychology, Harvard University, William James Hall, 33, Kirkland Street, Cambridge, MA 02138, USA. Email: [caram@wjh.harvard.edu](mailto:caram@wjh.harvard.edu)

The work reported here was supported in part by NIH grants NS22201 and DC0452 to Alfonso Caramazza. F.-Xavier Alario is supported by a post-doctoral grant from the Fondation pour la Recherche Médicale (ARI-2001110-9002/1). Albert Costa was supported by a post-doctoral fellowship from SISSA and a grant from the Spanish government (Programa Ramón y Cajal).

on-line speech production paradigms brings new methods to tackle issues that had previously been addressed within the speech-error tradition (Garrett, 1975). This approach should provide critical evidence on issues such as the organisation of syntactic and phonological encoding and the planning carried out by the speaker during the production of fluent speech.

In a recent study (Alario, Costa, & Caramazza, 2002) we examined the scope of encoding during speech production, that is to say: the amount of information that the speaker plans before he/she starts to produce an utterance. This issue was addressed with a classic psycholinguistic effect—the word frequency effect—in a noun phrase (NP) production task. In two experiments we asked participants to name pictures of coloured objects with simple NPs such as “the blue kite” that varied in frequency of the adjectives and the nouns. The analysis of the naming latencies showed additive effects of the frequency of the adjective and of the noun. This led us to the main conclusion of the paper: before articulation starts, the adjective and the noun are both processed at the level where frequency effects occur. For example, if we assumed that frequency affects the phonological encoding of the words, we would also have to assume that all the items in the NP undergo phonological encoding before articulation starts. This would go against the idea that only the first elements of the NP—the determiner and the adjective or, in other words, the first phonological word (see definition in Levelt, 2002)—are phonologically processed before utterance triggering.

Levelt (2002) criticises this study on two grounds, one theoretical and one empirical. First, he doubts that there is any theoretical proposal maintaining that phonological planning is limited to the first phonological word of an utterance. According to this view, the idea that only the first phonological word is encoded before articulation starts does not need to be refuted. Second, Levelt questions the validity of the interpretation of the frequency effect we report. He suggests that the observed effects could be due to a difference in recognition times for the pictures with low- and high-frequency names. In this commentary we address these two issues in turn. We will first discuss the assumptions that have previously been made about the unit and the scope of phonological encoding and show that the scope of phonological encoding is still an unresolved issue. Then we will discuss why the potential confound—e.g., speed of object recognition—does not provide a likely alternative account for the data we report.

## PLANNING DURING SPEECH PRODUCTION

The speech production system does not create and articulate utterances word by word, rather various (more than one) words are involved in different types of advance planning before articulation starts. At the same

time, it is commonly assumed that *whole* utterances (say, a sentence) do not need to be completely encoded before articulation starts. Processing of just part of the utterance at one representational level could trigger processing at the next level, so that different parts of the utterance are processed at different levels at the same time (the so-called incremental assumption; Bock and Levelt (1994); see e.g., Ferreira and Swets (2002) for discussion). Under these assumptions, it becomes necessary to specify how much information (e.g., how many words) has to have been processed to go from one level to the next one. In this commentary we will concentrate on the scope of processing at the level of phonological encoding.

Various authors have highlighted the central role of the phonological word in the process of phonological encoding. For example, in his proposal about phonological encoding, Levelt (1989) speculated about the scope of processing at this level and about the moment when the execution of a phonetic/articulatory plan can be triggered. He suggested that (p. 421) “execution can follow phonological encoding at a very short distance, a distance smaller than the phonological phrase. This distance is probably the size of a phonological word (the smallest ‘chunk’ delivered by the Phonological Encoder), and buffering will be minimal or absent”. In other words, the system would not need to wait until more than one phonological word is available before proceeding to the next (phonetic/articulatory) level. The assumption that the phonological word is the unit of phonological encoding during on-line speech production has been adopted by other authors. Wheeldon and Lahiri (1997) reported an on-line production experiment (Experiment 4) where participants produced simple sentences (e.g., the Dutch equivalent of “I drink John’s wine”). In this experiment, the authors found that the properties of the first phonological word—rather than the total number of phonological words in the sentence—were critical predictors of speech onset latencies. Wheeldon and Lahiri conclude that (p. 377) “this finding (note: of Experiment 4) provides strong evidence that the phonological word is the preferred unit of output in fluent speech production” (see also Wheeldon, 2000).

In principle, assuming that the phonological word is a critical unit of phonological encoding does not preclude the possibility that other larger phonological units might play a role during on-line utterance production. As Levelt (2002) points out (citing Meyer, 1996, and Schriefers & Teruel, 1999) the scope of phonological encoding could depend on different properties of the utterance being produced. In Levelt’s (1989) Chapter 10 the notion of the phonological phrase is underlined. Later, the following proposal is made (p. 420): “. . . phonological phrases are important units of phonological encoding. It is likely that the buffer is successively filled with phonological words, but that larger phonological phrase units are formed when the buffer is heavily loaded”. That is to say: when heavy processing

load situations occur, the scope of phonological encoding is made larger. The phonological encoder will then deliver more than one phonological word at a time to the next processing level. Wheeldon and Lahiri (1997) made a somewhat similar proposal, which they spell out in terms of the degree of incrementality used by the phonological encoder (p. 377): “the results of Experiment 4 do not rule out the possibility of nonincremental generation of prosodic structure. [...] It is therefore possible that with longer and more complex sentences the effects of whole sentence complexity may be observed in on-line sentence processing”.

These two proposals have in common the assumption that under *low load* conditions—which would probably need to be defined—phonological words will be delivered one by one to the articulator and that this delivery process could be different when longer or more complex utterances are produced. Consider the production of simple phrases or sentences in the light of these assumptions. If a speaker is asked to name pictures as fast as possible using NPs (e.g., “the blue kite”), he or she will be producing sequences involving *two phonological words*. This type of utterance is likely to involve the *smallest load* that an utterance comprising *more than one* phonological word can have. It certainly is no more complex than the utterances used by Wheeldon and Lahiri (1997). Therefore, under the assumptions reviewed earlier, it could be expected that the scope of phonological encoding will be minimal. During the production of these utterances, phonological words should be delivered one by one to the articulator.<sup>1</sup>

Aside from Wheeldon and Lahiri’s (1997) Experiment 4, other reports have also indicated that the scope of phonological encoding during the production of simple utterances is not larger than one phonological word. Schriefers and Teruel (1999) reported an NP production study and Meyer (1996) reported a study using simple conjunctions and simple sentences. In both studies, no phonological effects were found for items outside the first phonological word (see cited studies for details),<sup>2</sup> a result that supports the view that the unit of phonological encoding is not larger than one phonological word. However, other published results seem to indicate that during the production of *utterances as short and simple as NPs* the scope of

---

<sup>1</sup> The situation might be similar for the production of simple conjunctions such as ‘the arrow and the bag’ or sentences such as ‘the arrow is next to the bag’ (used by Meyer, 1996). However, this prediction is less straightforward than that made for NPs, since these utterances are somewhat more complex.

<sup>2</sup> In Meyer’s (1996) results, there was also a weak non-significant trend for a phonological effect for an item outside the first phonological word—the second noun of the utterances. This could be an indication that phonological encoding involves a unit larger than the first phonological word. Without more empirical evidence, it is difficult to ascertain the reliability of this observation.

phonological encoding could be larger than that. These results come either from picture-word interference experiments (Costa & Caramazza, 2002) or from determiner selection experiments (Alario & Caramazza, 2002; Miozzo & Caramazza, 1999), where phonological effects were found for items outside the first phonological word.

These apparently contradictory results, among other things, do not let us draw strong conclusions about the size of the phonological encoding carried out by the speaker before articulation starts. Clearly, more empirical evidence and theoretical considerations are required to solve the apparent contradictions between these two types of results and their interpretations (see, for example, the discussions in Alario et al., 2002, or in Costa & Caramazza, 2002). The study reported by Alario et al. (2002) was intended to provide evidence that would directly speak to this issue of planning during speech production.

It is certainly possible that the conclusions reached in studies of NP production (e.g., Alario et al., 2002) only apply to the production of this type of utterance. However, this does not undermine the purpose of the research. When conducting our experiments, we followed Meyer's recommendation (p. 480) "Even though speakers probably use different planning units in different situations, *it is important to find out which units they use in a given situation*" (emphasis added). This statement suggests that under particular "situations" phonological encoding could be conducted with fixed planning units. Whether this is true or not, it is important to describe the planning processes engaged during NP production, since NPs are one of the most commonly studied multi-word utterance types (since the seminal study by Schriefers, 1993).

Finally, Levelt (2002) notes that the study by Wheeldon and Lahiri (1997; see also Sternberg, Knoll, Monsell, & Wright, 1988) provides evidence that the number of phonological words in an utterance is a determinant of utterance onset latency. This could mean that utterances comprising various—up to three or four—phonological words can be completely planned in advance. Such a conclusion would go against the possibility of assuming a short scope of phonological encoding. Here a very important distinction must be made between two types of experimental situations: those where speech is delayed and initiated by a cue (Experiments 1 to 3 in Wheeldon and Lahiri (1997), and in Sternberg et al., 1988), and those where speech is produced on-line in a speeded fashion. In the first situation speakers are given ample time to prepare an utterance. Presumably they keep the prepared information ready in a short-term memory system. By asking participants to prepare utterances in advance, this paradigm critically *short-circuits the possibility of observing any incremental aspect of speech production*. Therefore, the results of this type of experiment, although interesting in their own right, cannot be used to

inform the nature of on-line speech encoding, which is the topic of Alario et al. (2002). This fundamental difference between the two types of situations is acknowledged by Wheeldon and Lahiri (1997). Levelt (2002) also notes it, but he still directly compares the results of these experiments with those of on-line production in his commentary.

In short, we have seen that *there are* theoretical proposals that underlie the role of phonological words as units of phonological encoding. These proposals suggest that phonological information is delivered phonological word by phonological word to the next processing level (phonetic/articulation), at least during the production of simple utterances and when “buffering is minimal”. Some empirical results seem to favour this view, while others suggest a larger scope of phonological encoding. In this context, our study was a motivated attempt to provide empirical evidence on the issue of planning during speech production using a simple utterance format.

What remains to be explained is why we think there are theories that assume that the phonological word can be the unit of encoding, while in his commentary Levelt asserts that “Alario et al. misinterpret the literature and fight a non-existing theoretical position”. In our view the disagreement stems from a different way of reading the existing positions in the literature. As we argued above, we based our study on proposals such as “phonetic spellout is probably made available to the articulator per phonological word” (p. 410), “as soon as all syllables for a phonological word have been planned, the Articulator can take over” (p. 411), “execution can follow phonological encoding at a very short distance, a distance smaller than a full phonological phrase. This distance is probably the size of a phonological word ...” (p. 421; all quotes taken from Levelt, 1989). Thus, we do not think we misinterpreted the theoretical positions, but rather we believe we tested some substantial claims that with more or less precision have been put forward in the literature.

## WORD-FREQUENCY EFFECTS IN PICTURE NAMING

The second criticism raised by Levelt (2002) is methodological in nature. In our experiments, we compared naming times for various groups of pictures. For example, we compared latencies for pictures representing objects with low- or with high-frequency names. We observed that the former are named slower than the latter, and we argued that the locus of this difference is at the level of lexical processing. Levelt (2002) points out that the difference in naming times could be attributable to other factors. For example if “pictures with low frequency names happen to be less recognisable than those with high frequency names, then any frequency

effect obtained in the experiment may just signal a visual process instead of a lexical one” (p. 668). In the following we argue that although this interpretation is logically possible, it is highly unlikely.

Various sources of evidence strongly suggest that the frequency effects we found are indeed attributable to lexical processing. There have been many studies that have manipulated word frequency in picture naming tasks and that have shown an effect of this variable on naming latencies (since the seminal study by Oldfield and Wingfield (1965)). Some of these studies also included control experiments of the type cited by Levelt (in press; e.g., picture categorisation tasks; for examples see Jescheniak & Levelt, 1994; Kroll & Potter, 1984; Levelt, Praamstra, Meyer, & Salmelin, 1998; Morrison, Ellis, & Quinlan, 1992; Wingfield, 1968). These control experiments were designed to assess whether the difference in naming performance between low- and high-frequency pictures was due to lexical access or to an earlier stage of processing. The rationale used is that categorising a picture requires processing at the early stages—visual processing and object identification—but it should not require name retrieval.

In the vast majority of cases, the manipulation of frequency that affected performance in the naming task did not affect performance in the categorisation task (Jescheniak & Levelt, 1994; Levelt et al., 1998; Morrison et al., 1992; Wingfield, 1968). As a matter of fact, we could find only one study reporting that a word frequency manipulation had an effect on picture recognition times, although only by-participant analysis was reported (Kroll & Potter, 1984). This widespread absence of word-frequency effects in picture categorisation tasks indicates that the manipulation of word-frequency affects the stage of lexical access.

As an example consider the study by Morrison et al. (1992). These authors report a picture naming experiment and a picture categorisation experiment—in the latter participants pressed one of two keys depending on whether the depicted object was *man-made* or *occurred naturally*. In the picture naming experiment Morrison et al. (1992) observed a clear effect of age of acquisition.<sup>3</sup> By contrast, there was clearly no effect of this variable (nor of word frequency) in the picture recognition experiment. Following the logic we have discussed, the authors interpreted this pattern of results as evidence that the effect of age of acquisition was truly attributable to the lexical retrieval stage of picture naming. The comparison of Morrison

---

<sup>3</sup> Recall that age of acquisition and frequency were confounded in our study and that we have been using the term “frequency effect” to refer to a phenomenon whose underlying cause could either be the frequency of the words or the age at which they are acquired. Here we will continue to treat frequency and age of acquisition as a single variable, under the provisions made in Alario et al. (2002).

et al.'s (1992) experiments with ours is particularly relevant because very similar picture stimuli were used in the two studies. These were simple black and white drawings with relatively high name agreement drawn from the norms developed in Snodgrass and Vanderwart (1980) or Cycowicz, Friedman, Rothstein, and Snodgrass (1997).<sup>4</sup> These pictures were created with the constraint that they represent *common*, easily identifiable objects in a very clear and simple fashion.

Another argument in favour of a lexical interpretation of the effect we report is the number of times that picture items were repeated in our experiments. Recall that in our study, pictures were presented in various colours: In Experiment 1, eight colours were used and in Experiment 2 four colours were used. Before the experiment proper, participants were always familiarised once with the materials. Therefore participants saw each object nine times in Experiment 1 and five times in Experiment 2. It has been reported that word frequency effects in picture naming are robust over number of repetitions. Jescheniak and Levelt (1994) reported that word frequency effects were robust even after pictures had been named three times. Levelt et al. (1998) showed that the effect was very much unchanged after as much as 12 repetitions of the items. These observations suggest that the core frequency effect is to a large extent insensitive to the number of repetitions. It is not clear that a potential effect affecting the visual recognition of the pictures would be stable over so many repetitions.<sup>5</sup>

Finally recall that as pointed out in Alario et al. (2002), clear frequency effects have been observed in the picture naming task for aphasic patients whose deficits could unambiguously be located at the level of word retrieval and not at the semantic/conceptual level of picture recognition or object identification (e.g., Caramazza & Hillis, 1990).

In summary then, the available empirical evidence points unambiguously to a lexical origin of the word frequency effects observed in the picture naming task. As we have seen, it is extremely unlikely that the difference in naming times between pictures with high- and low-frequency names is attributable to different object recognition speeds. Word frequency effects in an object recognition task—namely, object/non-object classification—have only been reported once. Therefore the results of the studies cited here and in Alario et al. (2002) provide grounds for assuming

---

<sup>4</sup> There were 28 pictures (out of 32) from these sources in our Experiment 1 and 44 (out of 50) in Experiment 2. The remaining pictures were created exactly in the same fashion as those already available.

<sup>5</sup> In the picture categorisation experiment where frequency effects were found (Kroll & Potter, 1984), participants were not familiarised with the materials: they saw the pictures only once during the experiment.

that the critical stage affected by the manipulation of word-frequency in the type of experiment that we have conducted is not object recognition but rather lexical retrieval.

## CONCLUSION

In this commentary we have addressed two criticisms raised in Levelt (2002) against the NP production study reported in Alario et al. (2002). The first criticism was that no available theoretical proposal stated that phonological planning is limited to the first phonological word of an utterance. In this paper, we have pointed to explicit proposals that make exactly this claim. Various authors have suggested that information could be delivered phonological word by phonological word to the phonetic/articulatory level of processing, at least in the type of production situation used in our experiments. This assumption, together with the seemingly conflicting results regarding how much phonological encoding is conducted before articulation starts, provided the motivation for our original study. The second criticism regarded the lack of a critical control in our experiments: the reported frequency effects could be attributable to object recognition speed rather than lexical access differences. A review of the available literature has shown that this claim is extremely unlikely given the experimental conditions used in picture naming experiments. In conclusion then, our study is theoretically well-motivated and the interpretation we report for our data remains valid.

Manuscript received February 2002

Revised manuscript received July 2002

## REFERENCES

- Alario, F.-X., & Caramazza, A. (2002). The production of determiners: Evidence from French. *Cognition*, *82*, 179–223.
- Alario, F.-X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, *17*, 299–319.
- Bock, K., & Levelt, W.J.M. (1994). Language production. Grammatical encoding. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego: Academic Press.
- Briggs, P., & Underwood, G. (1982). Phonological coding in good and poor readers. *Journal of Experimental Child Psychology*, *34*, 93–112.
- Caramazza, A., & Hillis, A.E. (1990). Where do semantic errors come from. *Cortex*, *26*, 95–122.
- Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding during speech production. *Journal of Memory and Language*, *46*, 153–177.

- Cycowicz, Y.M., Friedman, D., Rothstein, M., & Snodgrass, J.G. (1997). Picture naming by young children: norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*, 171–237.
- Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*, 57–84.
- Garrett, M.F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9). New York: Academic Press.
- Jescheniak, J.D., & Levelt, W.J.M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824–843.
- Kroll, J.F., & Potter, M.C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, *23*, 39–66.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W.J.M. (2002). Picture naming and word frequency. *Language and Cognitive Processes*, *17*, 663–671.
- Levelt, W.J.M., Praamstra, P., Meyer, A.S., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, *10*, 553–567.
- Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, *22*, 1–75.
- Lupker, S.J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory and Cognition*, *7*, 485–495.
- Lupker, S.J. (1982). The role of phonetic and orthographic similarity in picture-word interference. *Canadian Journal of Psychology*, *36*, 349–367.
- Meyer, A.S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, *35*, 477–496.
- Miozzo, M., & Caramazza, A. (1999). The selection of determiners in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 907–922.
- Morrison, C.M., Ellis, A.W., & Quinlan, P.T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory and Cognition*, *20*, 705–714.
- Oldfield, R.C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273–281.
- Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 841–850.
- Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology*, *11*, 17–50.
- Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.
- Sternberg, S., Kroll, R.L., Monsell, S., & Wright, C.E. (1988). Motor programs and the hierarchical organization in the control of rapid speech. *Phonetica*, *45*, 175–197.
- Wheeldon, L. (2000). Generating prosodic structure. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 249–274). Philadelphia: Psychology Press.
- Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*, 356–381.
- Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology*, *81*, 226–234.